

COMPARISON OF AUC_s OF BINORMAL ROC CURVE USING CONFIDENCE INTERVAL OF MEANS BASING ON 3 σ LIMITS

SURESH BABU

Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

ABSTRACT

In medical or health related problems, the phrase Receiver Operating Characteristic (ROC) curve plays a vital role to evaluate the assessment of classifier performance by its area under the curve (AUC). For binormal model, let the random variables (r.v's) X and Y be denoted by diseased normal population (D) and healthy normal population (H) with means μ_x and μ_y and sample variances s_x^2 and s_y^2 respectively. In binormal model area under the ROC curve lies between 0 and 1. This paper presents, an AUC which is to be prefixed so called Target AUC, (for example 0.9, 0.8 etc) to attain such AUC, 3 σ limits (i.e. σ , 2 σ and 3 σ) can be taken into the consideration to obtaining different AUCs basing on various possible distances (difference between population means) Δ_j , $j=1, 2, \dots, 9$. Nine possible distances can be calculated from lower and upper limits of the confidence interval of means which are based on 3 σ limits and also comparisons made between them. New average methods namely i) Simple Average Method ii) Fixed Weights Method (FW-Method) and iii) Proportional Weights Method (PW-Method) are briefly discussed also significant comparisons made among the three methods.

KEYWORDS: Comparison of AUC_s of Binormal ROC Curve Using, Radar Images, X and Y be Denoted by Diseased Normal Population (D) and Healthy Normal Population (H)

INTRODUCTION

The phrase Receiver Operation Characteristic (ROC) analysis had its origin in Statistical Decision Theory as well as Signal Detection Theory (SDT) and was used during II World War for the analysis of *radar images*. To draw a ROC curve, only the true positive rate (TPR) and false positive rate (FPR) are needed (as functions of some classifier parameters). The TPR defines how many correct positive results occur among all positive samples available during the test. FPR, on the other hand, defines how many incorrect positive results occur among all negative samples available during the test.

A ROC space is defined as TPR is plotted on the Y-axis and FPR is plotted on the X-axis. In other words, which depicts relative trade-offs between true positives and false positives Since TPR is equivalent with sensitivity and FPR is equal to 1 – specificity, the ROC graph is sometimes called the sensitivity verses (1 – specificity) plot. Each prediction result or instance of a confusion matrix represents one point in the ROC space.

The four possible states of the confusion matrix, which are non-negative integers shown below.

Table 1

Diagnosis	Test Result	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

The best possible prediction method would yield a point in the upper left corner or coordinate (0, 1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). The (0, 1) point is also called a *perfect classification*. In roc curve analysis, the classifier performance can often be explained by area under the roc curve (AUC). For a binary classifier it is shown that $AUC = (TP+TN) / (TP+FN+FP+TN)$. A poor classifier will have $AUC \leq 0.50$.

Earlier I have proposed new AUC estimates of the binormal roc model, using $100(1-\alpha)$ % confidence intervals and confidence interval of means basing on quartiles from 9 possible distances of two normal populations i.e. Diseased (D) and Healthy (H) populations with respective means and sample standard deviations.

In this paper, I have been observed 3σ limits (i.e. σ , 2σ and 3σ) as confidence limits of the confidence interval of means basing on 9 possible distances to the estimation of new AUCs. New summarized auc estimates have been derived by Simple Average Method, Fixed Weights Method (FW-Method) and Proportional Weights Method (PW-Method) and made comparisons among them also normality was tested by P-P plot.

AUC of the Binormal ROC Model

Let X and Y be two continuous random variables representing the test variables of Diseased (D) and Healthy (H) groups respectively, such that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$.

Then the parametric form of the binormal ROC is given by

$$ROC(t) = \Phi(a + b \Phi^{-1}(t)) \quad (1)$$

Where $a = \left(\frac{\mu_X - \mu_Y}{\sigma_X}\right)$ and $b = \left(\frac{\sigma_X}{\sigma_Y}\right)$ are constants.

Faraggi and Reiser (2002) have shown that AUC of the binormal ROC model, which is given by

$$AUC = \Phi\left(\frac{\mu_X - \mu_Y}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\right) \quad (2)$$

Let $\hat{\mu}_X$ and $\hat{\mu}_Y$ represent the estimated means and s_X^2 , s_Y^2 represent the sample variances of X and Y respectively. The estimated AUC can be obtained by replacing the parameters with their sample estimates. This gives

$$\widehat{AUC} = \Phi\left(\frac{\hat{\mu}_X - \hat{\mu}_Y}{\sqrt{s_X^2 + s_Y^2}}\right) \quad (3)$$

Suppose, we wish to develop an ROC curve such that it has a predetermined AUC denoted by AUC^* , like 0.9, 0.8 etc. If $Z^* = \Phi^{-1}(AUC^*)$ will be denotes the standard normal deviate corresponding to AUC^* . Faraggi and Reiser (2002) have shown that the estimated mean of the Diseased (D) group is

$$\hat{\mu}_X = \hat{\mu}_Y + Z^* \sqrt{s_X^2 + s_Y^2}. \quad (4)$$

With the help of $\hat{\mu}_X$, $\hat{\mu}_Y$ and s_X^2 , s_Y^2 , the AUC can be estimated.

Let Δ denotes the distance between the means of the two normal populations. Now Δ is defined as $\Delta = (\hat{\mu}_X - \hat{\mu}_Y)$

then the estimation of AUC can be reduced to the form of

$$AUC = \Phi \left(\frac{\Delta}{\sqrt{s_x^2 + s_y^2}} \right) \quad (5)$$

The true population mean of the two populations can be anywhere in the confidence interval (CI) the true difference between the means depending on the upper and lower limits of the intervals. In the following sections 3σ limits can be treated as upper and lower limits of the confidence interval of means.

The Possible Distances between the Means Basing on 3σ Limits

Usually the true mean of the population can be lie within the confidence interval. For each population there exist lower and upper limits of the confidence interval. For binormal distribution there exists 9 possible distances i.e. Δ_j ; $j=1, 2, \dots, 9$ which can be used to estimate new AUCs basing on 3σ limits (i.e. σ , 2σ and 3σ). Then equation (5) becomes

$$A_j = \Phi \left(\frac{\Delta_j}{\sqrt{s_x^2 + s_y^2}} \right); j=1, 2, \dots, 9 \quad (6)$$

Lower and upper limits of $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ limits are explained below.

Case I: Consider $\mu \pm \sigma$ limits

The Lower (L) and Upper (U) limits of $\mu \pm \sigma$ confidence interval are given by

Mean $\pm$ Standard Error of mean i.e. Mean $\pm \left(\frac{\sigma}{\sqrt{n}} \right)$

For diseased (D) normal population, the lower and upper limits of $\mu \pm 1\sigma$ are given by

Mean - $\left(\frac{\sigma_x}{\sqrt{n1}} \right) \leq \hat{\mu}_x \leq$ Mean - $\left(\frac{\sigma_x}{\sqrt{n1}} \right)$ i.e. $L2 = \hat{\mu}_x - \left(\frac{\sigma_x}{\sqrt{n1}} \right) \leq \hat{\mu}_x \leq \hat{\mu}_x - \left(\frac{\sigma_x}{\sqrt{n1}} \right) = U2$

For healthy (H) normal population, the lower and upper limits of $\mu \pm \sigma$ are given by

Mean - $\left(\frac{\sigma_y}{\sqrt{n2}} \right) \leq \hat{\mu}_y \leq$ Mean - $\left(\frac{\sigma_y}{\sqrt{n2}} \right)$ i.e. $L1 = \hat{\mu}_y - \left(\frac{\sigma_y}{\sqrt{n2}} \right) \leq \hat{\mu}_y \leq \hat{\mu}_y - \left(\frac{\sigma_y}{\sqrt{n2}} \right) = U1$

Case II: Consider $\mu \pm 2\sigma$ limits

The Lower (L) and Upper (U) limits of $\mu \pm 2\sigma$ confidence interval are given by

Mean $\pm 2 \times$ Standard Error of mean i.e. Mean $\pm 2 \left(\frac{\sigma}{\sqrt{n}} \right)$

For diseased (D) normal population, the lower and upper limits of $\mu \pm 2\sigma$ are given by

Mean - $2 \left(\frac{\sigma_x}{\sqrt{n1}} \right) \leq \hat{\mu}_x \leq$ Mean - $2 \left(\frac{\sigma_x}{\sqrt{n1}} \right)$ i.e. $L2 = \hat{\mu}_x - 2 \left(\frac{\sigma_x}{\sqrt{n1}} \right) \leq \hat{\mu}_x \leq \hat{\mu}_x - 2 \left(\frac{\sigma_x}{\sqrt{n1}} \right) = U2$

For healthy (H) normal population, the lower and upper limits of $\mu \pm 2\sigma$ are given by

Mean - $2 \left(\frac{\sigma_y}{\sqrt{n2}} \right) \leq \hat{\mu}_y \leq$ Mean - $2 \left(\frac{\sigma_y}{\sqrt{n2}} \right)$ i.e. $L1 = \hat{\mu}_y - 2 \left(\frac{\sigma_y}{\sqrt{n2}} \right) \leq \hat{\mu}_y \leq \hat{\mu}_y - 2 \left(\frac{\sigma_y}{\sqrt{n2}} \right) = U1$

Case III: Consider $\mu \pm 3\sigma$ limits

The Lower (L) and Upper (U) limits of $\mu \pm 3\sigma$ confidence interval are given by

$$\text{Mean} \pm 3 \times \text{Standard Error of mean i.e. Mean} \pm 3 \left(\frac{\sigma}{\sqrt{n}} \right)$$

For diseased (D) normal population, the lower and upper limits of $\mu \pm 3\sigma$ are given by

$$\text{Mean} - 3 \left(\frac{\sigma_x}{\sqrt{n_1}} \right) \leq \hat{\mu}_x \leq \text{Mean} + 3 \left(\frac{\sigma_x}{\sqrt{n_1}} \right) \text{ i.e. } L_2 = \hat{\mu}_x - 3 \left(\frac{\sigma_x}{\sqrt{n_1}} \right) \leq \hat{\mu}_x \leq \hat{\mu}_x + 3 \left(\frac{\sigma_x}{\sqrt{n_1}} \right) = U_2$$

For healthy (H) normal population, the lower and upper limits of $\mu \pm 3\sigma$ are given by

$$\text{Mean} - 3 \left(\frac{\sigma_y}{\sqrt{n_2}} \right) \leq \hat{\mu}_y \leq \text{Mean} + 3 \left(\frac{\sigma_y}{\sqrt{n_2}} \right) \text{ i.e. } L_1 = \hat{\mu}_y - 3 \left(\frac{\sigma_y}{\sqrt{n_2}} \right) \leq \hat{\mu}_y \leq \hat{\mu}_y + 3 \left(\frac{\sigma_y}{\sqrt{n_2}} \right) = U_1$$

The 9 possible true distances between the means basing on the limits are shown in Table 1.

Table 2: List of Possible Distances between the Means

Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9
$(L_2 - L_1)$	$(L_2 - \hat{\mu}_y)$	$(L_2 - U_1)$	$(\hat{\mu}_x - L_1)$	$(\hat{\mu}_x - \hat{\mu}_y)$	$(\hat{\mu}_x - U_1)$	$(U_2 - L_1)$	$(U_2 - \hat{\mu}_y)$	$(U_2 - U_1)$

Basing on these 9 possible distances different AUCs can be estimated for each case separately. New summarized AUC estimates can be obtained by three weighted average methods of all estimated AUCs are as given below.

- **Simple Average Method:** Among three methods, it is the simplest one. The estimation of new AUC through this method is given by

$$AUC_{AVG} = \sum_{j=1}^9 W_j A_j \text{ where } W_j = 1/9 \quad \forall j = 1, 2, 9$$

- **Fixed Weights Method (FW-Method):** In this method, AUC is defined as below

$$AUC_{FW} = \sum_{j \neq 5}^9 W_j A_j + \frac{0.5}{8} A_5 \text{ where } W_j = \frac{0.5}{8} \quad \forall j \neq 5$$

- **Proportional Weights Method (PW-Method):** In this method, AUC is given by the following formula

$$AUC_{PW} = \frac{\sum_{j=1}^9 W_j A_j}{\sum_{j=1}^9 W_j} \text{ where } W_j = \frac{1}{\Delta_j} \quad \forall j = 1, 2, 9.$$

Empirical Illustrations for Estimation of Binormal AUCs

In binormal ROC model, the true difference between the means of two normal populations attains merely the target AUC=0.9. To obtain more than one target AUC, different AUCs (A_j ; $j=1, 2 \dots 9$) can be estimated basing on 9 possible distances between the means of two normal populations i.e. Diseased (D) and Healthy (H) populations. These 9 possible distances can be obtained from each case of 3σ limits of the two normal populations such as $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ limits. For each case, new summarized improved AUCs can be estimated by three average methods basing on weights i.e. $\sum_{j=1}^9 W_j$ and to make comparisons among the three methods. Also normality was tested by P-P plot with respective R^2 value.

Illustration 1: Estimated AUC values and new summarized AUC estimates with normality fit by P-P plot from $\mu \pm \sigma$ limits with R^2 value are shown in the following Table 2.

Table 3: Estimated AUC Values and Normality Fit by P-p Plot from $\mu \pm \sigma$ Limits

Target AUC = 0.9, $\mu_Y = 85$			
S. No	Inputs	Estimated AUC values	Normality Fit by P-P Plot
1	$n_1 = 5, n_2 = 5$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_5\} = \{0.9000, 0.8328, 0.7419, 0.9450, 0.9000\}$, $\{A_6..A_9\} = \{0.8328, 0.9722, 0.9450, 0.9000\}$ $AUC_{AVG} = 0.8855, AUC_{FW} = 0.8918, AUC_{PW} = 0.8626$	$Y = 0.069X + 0.885$ $R^2 = 0.907$
2	$n_1 = 10, n_2 = 15$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_5\} = \{0.8926, 0.8550, 0.8093, 0.9284, 0.9000\}$ $\{A_6..A_9\} = \{0.8641, 0.9543, 0.9339, 0.9070\}$ $AUC_{AVG} = 0.8938, AUC_{FW} = 0.8965, AUC_{PW} = 0.8855$	$Y = 0.044X + 0.893$ $R^2 = 0.964$

Illustration 2: Estimated AUC values and new summarized AUC estimates with normality fit by P-P plot from $\mu \pm 2\sigma$ limits are shown in the following Table 3.

Table 4: Estimated AUC Values and Normality Fit by P-P Plot from $\mu \pm 2\sigma$ Limits

Target AUC = 0.9, $\mu_Y = 85$			
S. No	Inputs	Estimated AUC values	Normality Fit by P-P Plot
1	$n_1 = 5, n_2 = 5$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_5\} = \{0.9000, 0., 0.7419, 0.5066, 0.9722, 0.9000\}$, $\{A_6..A_9\} = \{0.7419, 0.9946, 0.9722, 0.9000\}$ $AUC_{AVG} = 0.8477, AUC_{FW} = 0.8706, AUC_{PW} = 0.5413$	$Y = 0.145X + 0.847$ $R^2 = 0.829$
2	$n_1 = 10, n_2 = 15$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_5\} = \{0.8848, 0.7980, 0.6805, 0.9502, 0.9000\}$ $\{A_6..A_9\} = \{0.8203, 0.9819, 0.9581, 0.9137\}$ $AUC_{AVG} = 0.8764, AUC_{FW} = 0.8867, AUC_{PW} = 0.8316$	$Y = 0.092X + 0.876$ $R^2 = 0.910$
3	$n_1 = 25, n_2 = 20$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_4\} = \{0.9057, 0.8410, 0.7525, 0.9450, 0.9000\}$ $\{A_6..A_9\} = \{0.8328, 0.9700, 0.9411, 0.8940\}$ $AUC_{AVG} = 0.8869, AUC_{FW} = 0.8926, AUC_{PW} = 0.8668$	$Y = 0.066X + 0.886$ $R^2 = 0.929$

Illustration 3: Estimated AUC values and new summarized AUC estimates along with normality fit by P-P plot with R^2 value from $\mu \pm 3\sigma$ are shown in the following Table 4.

Table 5: Estimated AUC Values and Normality Fit by P-P Plot from $\mu \pm 3\sigma$ Limits

S. No	Inputs	Estimated AUC values	Normality Fit by P-P Plot
1	$n_1 = 5, n_2 = 5$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_5\} = \{0.9000, 0., 0.6304, 0.5000, 0.9871, 0.9000\}$, $\{A_6..A_9\} = \{0.6304, 0.9993, 0.9871, 0.9000\}$ $AUC_{AVG} = 0.8260, AUC_{FW} = 0.8584, AUC_{PW} = 0.7420$	$Y = 0.173X + 0.826$ $R^2 = 0.831$
2	$n_1 = 10, n_2 = 15$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_5\} = \{0.8767, 0.7293, 0.52515, 0.9663, 0.9000\}$ $\{A_6..A_9\} = \{0.7685, 0.9938, 0.9746, 0.9199\}$ $AUC_{AVG} = 0.8505, AUC_{FW} = 0.8721, AUC_{PW} = 0.6222$	$Y = 0.142X + 0.850$ $R^2 = 0.859$
3	$n_1 = 25, n_2 = 20$ $\mu_x = 103.12$ $s_x = 10, s_y = 10$	$\{A_1..A_4\} = \{0.9085, 0.8044, 0.6491, 0.9604, 0.9000\}$ $\{A_6..A_9\} = \{0.7902, 0.9854, 0.9560, 0.8909\}$ $AUC_{AVG} = 0.8717, AUC_{FW} = 0.8841, AUC_{PW} = 0.8112$	$Y = 0.101X + 0.871$ $R^2 = 0.888$

Comparison of New Summarized AUC Estimates

After obtaining estimated AUC values basing on 9 possible distances from each case, new summarized AUC estimates can be computed by three average methods such as Simple Average Method, Fixed Weights Method (FW-Method) and Proportional Weights Method (PW-Method) and also significant comparisons made among the three methods are shown in the following diagrams.

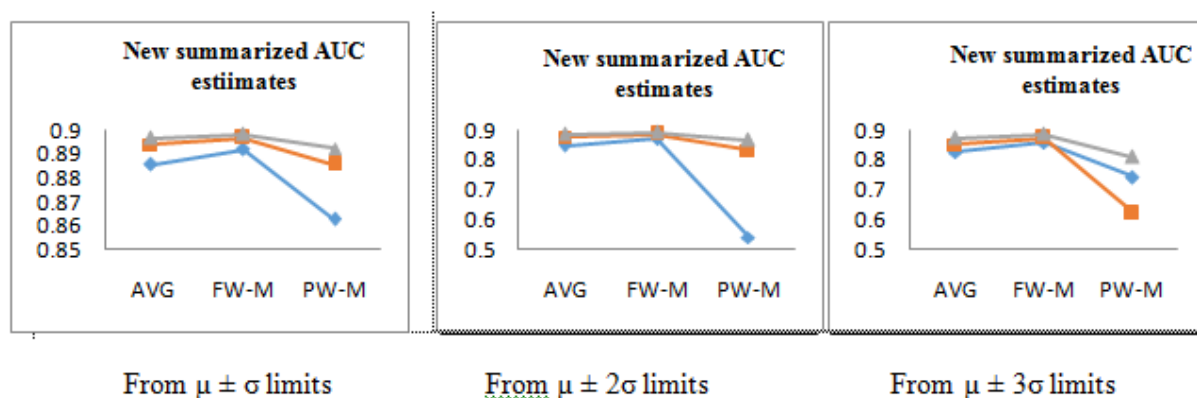


Figure 1

CONCLUSIONS

At fixed mean of healthy normal population $\mu_Y = 85$, diseased normal population mean has been estimated as $\mu_x = 103.12$. At both μ_x , μ_Y and sample standard deviations S_x , S_y , different AUC estimates have been obtained using confidence interval of means basing on 9 possible distances from the confidence limits which can be obtained from 3σ limits i.e. $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ limits. Among all three cases, in each case middle most possible distance used AUC would attain the Target AUC=0.9 and also observed that R^2 value has been extensively increased as sample size increases. In three cases, while considering the $\mu \pm \sigma$ limits, the estimated AUC values, new summarized AUC estimates and R^2 values have been given an appropriate values and found that consideration of $\mu \pm \sigma$ limits would give best results among the three. Among three average methods, Fixed Weight Method provides better summarized AUC estimates other than two methods. And finally normality was tested by P-P plot which shows R^2 values are very close to 1 in case I. Moreover consideration of $\mu \pm \sigma$ limits is the best one as compared to $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ limits.

REFERENCES

1. Metz, C. E., 1978, "Basic Principles of ROC analysis", Seminars in Nuclear Medicine, 8, pp. 283 – 298.
2. Hanley, A., James, and Barbara, J. Mc., Neil, 1982, "A Meaning and Use of the area under a Receiver Operating Characteristic (ROC) curves", Radiology, 143, pp. 29 - 36.
3. Lloyd, C. J., 1998, "The Use of smoothed ROC curves to summarize and compare diagnostic systems", Journal of American Statistical Association, 93, pp. 1356 – 1364.
4. Metz et al., 1998, "Maximum Likelihood Estimation of receiver operating characteristic (ROC) curves continuously distributed data", Statistics in Medicine, 17, pp. 1033-1053.
5. Pepe, MS., 2000, "Receiver operating characteristic methodology", Journal of American Statistical Association, 95, pp. 308 – 311.
6. Farragi David, and Benjamin Reiser, 2002, "Estimation of the area under the ROC curve, Statistic in Medicine", 21, pp. 3093 - 3106.
7. Lasko, T. A., 2005, "The use of receiver operating characteristics curves in biomedical informatics", Journal of Biomedical Informatics, vol. 38, pp. 404 –415.

8. Krzanowski, J. Wojtek., and David, J. Hand., 2009, "ROC Curves for Continuous Data".
9. Sarma K.V.S., et., al 2010, "Estimation of the Area under the ROC Curve using Confidence Intervals of Mean", ANU Journal of Physical Sciences.
10. Suresh Babu.N. and Sarma, K. V. S., 2011. "On the Comparison of Binormal ROC Curves using Confidence Interval of Means", International Journal of Statistics and Analysis, Vol. 1, pp. 213-221.
11. Suresh Babu. N and Sarma K.V.S. (2013), On the Estimation of Area under Binormal ROC curve using Confidence Intervals for Means and Variances. International Journal of Statistics and Systems, Volume 8, Number 3 (2013), pp. 251-260.
12. Suresh Babu. N and Sarma K. V. S. (2013), Estimation of AUC of the Binormal ROC curve using confidence intervals of Variances. International Journal of Advanced Computer and Mathematical Sciences, Volume 4, issue 4, (2013), pp. 285-294.
13. Suresh Babu. Nellore (2015), Area Under The Binormal Roc Curve Using Confidence Interval Of Means Basing On Quartiles. International Journal OF Engineering Sciences & Management Research. ISSN 2349-6193, Impact Factor (PIF): 2.243.
14. Suresh Babu. Nellore (2015), Various performance measures in Binary classification—An Overview of ROC study. International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015. ISSN 2348-7968, International Impact Factor: 1.5

